USING A GENERATIVE ADVERSARIAL NETWORK TO EXPLORE THE NEW AESTHETIC

Angela May Xie

CS 379H
Department of Computer Science
The University of Texas at Austin

TC 660H
Plan II Honors Program
The University of Texas at Austin

May 15, 2017

Donald S. Fussell, Ph.D.
Department of Computer Science
Supervising Professor

Bogdan Perzyński, M.F.A.
Department of Art and Art History
Second Reader

ABSTRACT

Author: Angela May Xie

Title: Using a Generative Adversarial Network to Explore the New Aesthetic

Supervising Professor: Donald S. Fussell

The purpose of this thesis was to explore ways of creating computer art using an activation maximization generative adversarial network (AM GAN). GANs are a recent development in machine learning. AM GANs, in particular, use the GAN model to generate images that highly activate a specific neuron within an image recognition neural network. I frame my project in the context of New Aesthetic, an art movement that focuses on the collaboration between humans and digital technology. Since its emergence, New Aesthetic has been criticized for a few different reasons. I wanted to address these criticisms using an AM GAN while also exploring ways that an AM GAN can be controlled.

TABLE OF CONTENTS

INTRODUCTION

For my thesis, I wanted to do a project on computer art, specifically art generated from artificial intelligence (AI). The first landmark AI art project was Harold Cohen's AARON, which Cohen started developing AARON in 1973 and continued to do so until his death in 2016. To summarize, AARON was an image-generating computer program that used a state machine programmed by Cohen to determine what is drawn next based on what had been drawn previously.[1] At the time, most artificial intelligence software was similar to AARON: manually hard-coded to complete a specific task. But since as we enter the 21st century, AI implemented through machine learning has become more prevalent and more preferred. For my own project, I wanted to take advantage of new technologies in machine learning to create computer art and came across activation maximization generative adversarial networks (AM GANs), which utilize deep learning to create images based on what an image recognition neural network has learned.

In addition to using new technology, I wanted to frame any art generated in the context of contemporary art. I chose to study New Aesthetic, a recent art movement focused the collaboration between humans and digital technology. As computers have yet to become sentient on their own, computer art by nature is art created from the collaboration between humans and technology and is thus included in the definition of New Aesthetic. New Aesthetic has received a lot of criticism regarding what it can achieve, and I wanted to use an AM GAN to address these criticisms.

The goal of my thesis is to provide an example of how computer art can counter criticisms of New Aesthetic while also exploring in what ways AM GANs can be controlled to

[1] Harold Cohen, Becky Cohen, and Penny Nii, *The First Artificial Intelligence Coloring Book* (Los Altos, CA: W. Kaufmann, 1984).

create art. This thesis text will be organized in the following way. The Background section will present criticisms of New Aesthetic and will argue how using an AM GAN to generate images addresses these criticisms. Looking at images generated by an AM GAN in the context of New Aesthetic provides grounding for the artistic application of this technology. The Work section will outline in which ways aspects of the AM GAN were explored to create images that fit into the New Aesthetic.

BACKGROUND

New Aesthetic

In our current times, we are more reliant on digital technology than ever, and the New Aesthetic articulates how this collaboration with technology has changed the way we perceive the world. The term "New Aesthetic" was coined by London-based artist James Bridle in his tumblr microblog "The New Aesthetic" and gained traction after its debut during a 2012 South by Southwest (SXSW) panel lead by Bridle.[2] Bridle particularly wanted to focus on the dissolve of lines between the physical and virtual, the real and the digital.[3]

This dissolve has lead to human experiences that did not exist before. For example, what did it mean thirty years ago to "like" someone's picture on social media? Nothing. But as digital technology in the form of media becomes commonplace, what is common to human experience must change as well and we ought to create new ways of communicating such phenomena. Bridle himself coined the term *Strasseblickfernweh*, Street View Wanderlust, to describe that "when you see a distant place through the internet and a number of devices… and wish you were there."[4] The experience Bridle is describing may be oddly specific, but the Street View feature is something most of us are familiar with that has certainly changed how we are able to perceive physical spaces and is a piece of digital technology that certainly did not exist before the boom of digital media.

For the most part, art is meant to reflect human experiences, and the New Aesthetic is no different. As implied by its name, the New Aesthetic is meant to be an art movement. New

[2] Ian Bogost, "The New Aesthetic Needs to Get Weirder," The Atlantic, April 13, 2012.
[3] James Bridle, "Sxaesthetic," Booktwo, March 15, 2012.
[4] Bridle, "Sxaesthetic."

Aesthetic art is the visual representation of how the New Aesthetic manifests. Examples of what is considered New Aesthetic vary greatly but share the common properties of being, in some way, computer-generated or heavily inspired by digital imagery. This includes anything data visualization to glitch art to enlarged pixels (think 8-bit art).[5]

As with any new self-proclaimed movement, the New Aesthetic was both praised and criticized shortly after its debut at SXSW. Bruce Sterling's "An Essay on the New Aesthetic" noted positive aspects of the New Aesthetic as well as some critical flaws. Ian Bogost's "The New Aesthetic Needs to Get Weirder" drew upon Sterling's essay to further point out the New Aesthetic's limits.

The most positive and the most important aspect of the New Aesthetic is that it is truthful. Referring to the images collected on "The New Aesthetic" tumblr, Sterling points out that "scarcely one of the real things in there would have made any sense to anyone in 1982, or even in 1992." The New Aesthetic accurately reflects the state of the modern world. The digital technology in the early 21st century is very different from that of the late 20th century. What is so real and integral to our everyday experiences is completely alien and surreal to our past selves twenty-some years ago. Furthermore, the New Aesthetic reflects a sort of cultural agnosticism created by the ubiquity of the internet. Though the New Aesthetic has its origins in London, its images are not foreign to people halfway around the globe since the internet and social media culture do not vary much in varying locations.[6]

The main criticism of the New Aesthetic is that its definition do not appear to be thought about critically by those who are a part of it. Sterling enumerates all the ways in which the digital

---

[5] Bruce Sterling, "An Essay on the New Aesthetic," Wired, April 02, 2012.
[6] Sterling, "An Essay on the New Aesthetic."

technology represented in New Aesthetic images are not very advanced or even new, thus revealing a shallow understanding of technology. Furthermore, New Aesthetic images vary so drastically in content that it is difficult to clearly define what constitutes New Aesthetic. As a result, New Aesthetic as an artform can appear lazy. For example, New Aesthetic artists seem to interpret glitches and corruption artifacts in rendered images as flaws in computer vision yet Sterling points out that glitch art images indicate a failure in machine processing and computer displays.[7] While Bridle rejoices in "#sxaesthetic" at the idea that "there [are] new and extraordinary things and experiences in this world, like the ability to *see through satellites*," Sterling denies the newness and extraordinariness of satellite images, dating them back to the Aero-Futurism movement of the 1930s and its failure due to boringness.[8] [9] Lastly, while Sterling does acknowledge how 8-bit art, particularly sculpture and architecture, is successful in breaking down a barrier between the digital and physical, he ultimately calls the allusion to '80s graphics "cute" "sentimental fluff."[10] In summary, Sterling sees the biggest problem with the New Aesthetic is that it is trying "to hack a modern aesthetic, instead of thinking hard enough and working hard enough to build one," yet he believes it has a lot of potential to grow.[11] As evidenced by "The New Aesthetic" microblog, much of New Aesthetic art at the moment relies on collecting screenshots, curating rather than creating. As noted by Sterling, the curation of New Aesthetic appears to be rather indiscriminate. There is a lot of room for artists to be creative

[7] Sterling, "An Essay on the New Aesthetic."
[8] Bridle, "Sxaesthetic."
[9] Sterling, "An Essay on the New Aesthetic."
[10] Sterling, "An Essay on the New Aesthetic."
[11] Sterling, "An Essay on the New Aesthetic."

within the realms of New Aesthetic — after all, digital technology allows people to make things previously unthinkable — but they have yet to do so.

Like Sterling, Bogost believes the New Aesthetic can be more interesting, but in contrast to Sterling's criticism on the lack of limits on the New Aesthetic, Bogost finds the New Aesthetic too focused on human experiences. Bogost is a part of a philosophical movement called Object Oriented Ontology (OOO), which rejects the idea that humans ought to be the focus of ontology, the study of existence. According to Bogost, the New Aesthetic has acknowledged that computers "have taken on lives of their own," and this allows for art that is not human-centric. However, New Aesthetic thus far "is still primarily interested in human experience," limiting its images to that of computational media. In this aspect, Bogost's criticism recalls Sterling's in that the computational technology explored by the New Aesthetic is not very new or advanced, just more prevalent to the human experience.[12]

Technology as a Hyperobject

In order to address Bogost's criticism that the New Aesthetic has the potential to explore Object Oriented Ontology yet chooses not to, I need to know how. I looked to Timothy Morton's *Hyperobjects*, a book that discusses the role of hyperobjects in OOO and the human world, including how humans can interact with hyperobjects through art. Morton defines a hyperobject as a "things that are massively distributed in time and space relative to humans" and are "hyper in relation to some other entity." Hyperobjects share a series of similar properties: viscosity, nonlocality, temporal undulation, and phasing.[13] In this section, I will argue why digital technology could be considered a hyperobject so the ways Morton prescribes for interacting with

[12] Bogost, "The New Aesthetic Needs to Get Weirder."
[13] Timothy Morton, *Hyperobjects* (Minneapolis: University of Minnesota Press, 2014), 1.

hyperobjects can be applied to technology and used to address Bogost's criticism of New Aesthetic.

Viscosity entails that the hyperobject dissolves the boundaries of objects it surrounds.[14] As evidenced in recent years, digital technology has begun dissolving the boundaries of what it means to be humans. Many of us have lives that is dependent upon digital technology. We set alarms on the little computers we call cell phones to wake us up in the morning. We do almost all our work, receive almost all our information, whether it be through news sites or online interactions with others, through our computers. To a lot of people, their cell phone or personal computer is almost like an appendage. Some people are even going as far as implanting computer chips into their hands in order to have access to their digital information at all times.[15]

Nonlocality means that a hyperobject is exists in many spaces simultaneously and that we can only observe local manifestations of it, which are not directly the hyperobject.[16] Digital technology's nonlocality is exemplified through our inability to point to a quintessential example of digital technology. What exactly is technology? Is it the physical computers we all have in our bags and pockets? Is it the software — operating systems, programs, apps — that give us an interface for interacting with our digital devices? Or is it the network of all such devices, software, and the data they contain and produce, i.e., the internet? All the those this listed have properties that other examples of technology do not and are thus simply manifestations of digital technology. Digital technology itself is too large for us to pin down to an object.

---

[14] Morton, *Hyperobjects*, 30.
[15] Charlie Warzel, "Satan's Credit Card" BuzzFeed, May 21, 2016.
[16] Morton, *Hyperobjects*, 1.

Temporal undulation is the property that a hyperobject exists for an extremely long time. The existence of a hyperobject is not infinite but is so unfathomably long that it may as well seem infinite.[17] This is where the argument for technology as a hyperobject may fail. In the scope of the history of the universe, or even just the short history of human existence, digital technology has been around for only a minuscule fraction of time: a century at most. However, one can argue that it's likely that digital technology is here to stay, that it has acquired a permanent role in human lives and will continue to exist beyond the anthropocene, though in a manifestation that is different from how we perceive technology today.

Phasing is the result of the temporal and spatial vastness of a hyperobject as the result of temporal undulation and nonlocality and entails that we can only observe the small intersection between our plane of existence and that of the hyperobject.[18] Not only are there varying manifestations of digital technology existing at a specific time, the manifestations of digital technology has varied over the timeline of its existence. For example, manifestations of digital technology that were common a decades ago, for example, floppy discs, rarely exist today except for novelty purposes.

In the final chapter of *Hyperobjects*, Morton describes how we can collaborate with hyperobjects to create art. Morton defines art as "a conversation between what we think we know and what materials we have at our disposal," and as we learn more about hyperobjects, art "strives to attune itself to hyperobjectivity."In our current time, in which we begin privileging the role of hyperobjects, art "must be a *tuning* to the object" and "becomes a collaboration between humans and nonhumans." New Aesthetic is evidence of the collaboration between

[17] Morton, *Hyperobjects*, 60.
[18] Morton, *Hyperobjects*, 70.

humans and digital technology. However, it does not necessarily *tune* to the object. Tuning to an object, like tuning an instrument, requires deeper than surface level understanding of an object. As revealed in Sterling's criticism of the New Aesthetic, many New Aesthetic artists fail to show advanced knowledge of technology and therefore cannot successfully tune to technology. Throughout *Hyperobjects*, Morton notes that "the more we know about an object, the stranger it becomes."[19] As the object becomes more strange to us, we allow it to do more work in our collaborations with it. We know enough about it to have a degree of control over it while also trusting it enough to operate on its own. The New Aesthetic that addresses OOO finds a balance between having control over technology and letting it do its thing.

Using a Generative Adversarial Network to Create Images

In recent years, generative adversarial networks (GANs) have been an up-and-coming area in the machine learning field within computer science. The concept of GANs was created by Ian Goodfellow et al. and first published in 2014. A generative adversarial network is described as a deep-learning framework that is composed of two models trained simultaneously: a generative model and a discriminative, or adversarial, model. The goal of the generative model is to generate some sort of output that mimics the training data while the goal of the adversarial model is to determine the likelihood that the generated output is part of the training data, assigning it a "score." The generative model then uses the feedback to attempt to generate better output. This back and forth between the two models make up an iteration, and after many iterations, the generated output eventually converges. To better explain GANs, the original paper compared the generative model to counterfeiters, who are attempting to produce fake money, and

[19] Morton, *Hyperobjects*, 161-175.

the discriminative model to the police, who are trying to detect fraud. The counterfeiters are successful when what they produce is indistinguishable from the "real" thing.[20]

GANs come with a variety of advantages over previous generative machine learning systems. The most notable advantage is that the framework can be used on a wide variety of model and data types. Furthermore, the GAN allows for data distributions that are sharp or even degenerate, while a generative framework such as Markov chain would require data distributions that are at least somewhat blurry. GAN training is also unsupervised, meaning it is not necessary to make inferences regarding the training data.[21]

Anh Nguyen et al. created and published in 2016 work on using a GAN create a method called activation maximization (AM) that synthesizes an input that highly activates a neuron in a deep neural network. In particular, Nguyen et al.'s work on activation maximization focused on synthesizing an image that would highly activate a chosen neuron in an image recognition deep neural network (DNN), in particular, CaffeNet. Within an image recognition DNN are series of neurons that each representing a particular object, ending with a neuron that outputs the probability that the particular object is present in an image. This final probability neuron is what the AM GAN is trying to highly activate. AM thus allows for a visualization of what a neuron has learned to be a canonical representation of a given object. In the context of the GAN framework, a deep generator network serves as the generative model and synthesizes an image, which is then rated using the output of the chosen probability neuron, the adversarial model. With each iteration, the generative model modifies the previously generated image in an attempt to receive a higher score. An initial starting image can be set to a particular image in order to

[20] Ian Goodfellow et al., "Generative Adversarial Networks," July 10, 2014, ArXiv.
[21] Goodfellow et al., "Generative Adversarial Networks."

influence the outcome of the synthesized image, or it can be set to a randomly generated image of noise in order to produce an original image.[22]

The following equation represents what AM is trying to solve:

$$\hat{y} = \arg\max_{y}(\Phi_h(G(y)) - \lambda|y|)$$

Here, y is a code representation of an image such that G(y) is an image generated by deep generator network that highly activates target neuron h in DNN Φ. $\Phi_h(G(y))$ is thus the probability that G(y) is an image containing object h. λ is a L_2 regularization factor that was found empirically by Nguyen et al. to be 0.005. AM thus tries to find a code representation of an image that best activates target neuron h in DNN Φ.[23]

The same AM technique can also be used to find an image that best activates two target neurons, $h1$ and $h2$, which is represented in the following equation:

$$\hat{y} = \arg\max_{y}(\Phi_{h1}(G(y)) + \Phi_{h2}(G(y)) - \lambda|y| - \gamma\left|\Phi_{h1}(G(y)) - \Phi_{h2}(G(y))\right|)$$

In words, this means that AM tries to find a code representation of an image that highly activates target neurons $h1$ and $h2$ while maintaining that one neuron does not dominate the other. To ensure the latter condition, a penalty factor γ, which is found empirically, is used to make sure both target neurons are equally activated. While the output images generated by activating two neurons at once are not realistically representational of any particular objects, they are described as “artistically interesting.”[24]

[22] Anh Nguyen et al., "Synthesizing the preferred inputs for neurons in neural networks via deep generator networks," May 30, 2016, ArXiv.
[23] Nguyen et al., "Synthesizing the preferred inputs.”
[24] Nguyen et al., "Synthesizing the preferred inputs.”

Nguyen et al.'s activation maximization system could be used to generate images that not only fit into the New Aesthetic but also address some of its criticisms. An AM GAN embodies the New Aesthetic in that it takes in human-provided training data but learns in an unsupervised fashion. Much of the New Aesthetic is about how technology has changed the way we perceive our world. The AM GAN is able to generate canonical representations of objects, indicating that our representation and thus perception of objects through images have changed due to the prevalence of digital technology. The AM GAN provides a sort of conceptualization of how objects are represented in digital images; the images generated by the AM GANs are visual summaries for images of particular objects. What distinguishes saving images generated from a GAN from simply taking screenshots is that GANs can be programmed, trained, manipulated, and this allows for the ability for New Aesthetic artists to create rather than curate. Furthermore, GANs are very new — in fact, the concept of GANs is more recent than that of the New Aesthetic — thus addressing the criticism that many New Aesthetic images do not depend on computer technologies that truly groundbreaking.

Generating images using an AM GAN is also consistent with object oriented ontology. First, it acknowledges that a computer, an artificial intelligence, has learned representations of other objects through images in a way that is different from what humans know. Second, creating art using an AM GAN allows for the collaboration between humans and an hyperobject, as prescribed by Morton. There are ways for humans to influence the output of the AM GAN by selecting different inputs, but ultimately, it is the computer that creates the images.

WORK

One of the goals of my thesis has been to identify and explore ways an activation maximization generative adversarial network can be used to create images that fit into the New Aesthetic while addressing its criticisms. The AM GAN requires an image recognition DNN as one of its inputs, and throughout this project, I used CaffeNet, because that was what Nguyen et al. had used.

I was able to identify at least five different ways in which AM GANs can be influenced to create different desired outputs. There are as follows:

- Selection of neurons. This includes both which particular target neurons are selected and the number of neurons selected.
- The initial image used as input.
- The number of iterations the GAN goes through.
- Modifications to the image in between iterations. This could mean applying an image kernel that alters the color (for example, removing all pixels of a certain color) or geometry (for example, a blur or warp effect) of an image.
- Image recognition DNN used.

For the purpose of this thesis, I chose to focus on the first three ways.

Selecting Three Neurons at Once

Nguyen et al. had mentioned the artistic potential of generated images that maximally activated two target neurons at once so it seemed like a natural next step to try to generate images that maximally activated three target neurons. In this section, I will describe my approach in formulating an AM method for generating images that highly activated three target neurons simultaneously.

Recall that the AM equation for generating an image that highly activates two target neurons is as follows:

$$\hat{y} = \arg\max_{y}(\Phi_{h1}(G(y)) + \Phi_{h2}(G(y)) - \lambda|y| - \gamma\left|\Phi_{h1}(G(y)) - \Phi_{h2}(G(y))\right|)$$

Thus, it follows that the AM equation for generating an image that highly activates three target neurons would be:

$$\hat{y} = \arg\max_{y}(\Phi_{h1}(G(y)) + \Phi_{h2}(G(y)) + \Phi_{h3}(G(y)) - \lambda|y| - p)$$

$$p = \gamma(\left|\Phi_{h1}(G(y)) - \Phi_{h2}(G(y))\right| + \left|\Phi_{h1}(G(y)) - \Phi_{h3}(G(y))\right| + \left|\Phi_{h2}(G(y)) - \Phi_{h3}(G(y))\right|)$$

p multiplies the penalty factor γ with all pairwise differences of how activated each target neuron is.

In order to determine the best value for γ, I generated images using different values of γ in between 0 and 1, first in step sizes of 0.1 and then in step sizes of 0.01 after narrowing down the range. A value of γ was determined to be "better" if it was more obvious in the output image what target neurons were being activated. For all images generated, I using the same set of three neurons (lamp, lemon, and candle) and allowed the GAN to run for 300 iterations.

However, I ultimately determined that the best value for γ is zero or very close to zero. This is the result of a simple combinatorics problem: as the number of target neurons increases linearly, the number of pairwise differences increases exponentially. When the cumulative differences between how highly activated each target neuron is grows too large, $\hat{y}$ can result in a junk image in which none of the target neurons are that activated. Even though zero is empirically the best value for γ, setting γ to zero during AM runs the risk of one neuron dominating the resulting image or one neuron having less of an effect.

Figure 1: The image on the left was generated using the target neurons {broom, balloon, lemon}. The image on the right was generated the target neurons {lemon, candle, lamp}.

Figure 1 shows two different images generated from AM using three target neurons. In both images, there seems to be one target neuron that dominates the final output image or a target neuron that is less present than the others. The image on the left clearly shows several brooms. Effects of the other two target neurons are in fact present but are less obvious: presence of the hot air balloon neuron is indicated by bright streaks of red, blue, and yellow, and the overall geometric composition of the image resembles the cross-section of a lemon. The image on the right shows a lamp with qualities of a candlestick; the lemon neuron is only vaguely represented through the bright yellow lampshade.

The significance of generating images that activates three target neurons simultaneously is similar to that of images that activate two neurons at once. While these images do not necessarily represent realistic objects, the generation of such images allows for a formal method of visualizing the combinations of three objects. Typically, such combinations are considered abstract and imagining them is difficult for humans (though possible in some cases). However, with an AM GAN, the task becomes easily achievable for a computer.

Effect of Initial Image

I tested the effect of an initial image on the generated output image by setting the initial to "checkerboard" images of different colors. I used an image of randomly-generated noise as a control. During these experiments, the GAN ran for 300 iterations.

Figure 2: These are all images generated with the target neuron set to balloon. The image on the left was not generated using a specified initial image. The image in the center was generated using an initial image that was a 5x5 green and red checkerboard pattern. The image on the right was generated using an initial image that was a 5x5 black and white checkerboard pattern.

Figure 2 shows the results of one of these experiments. As shown, the initial image does have an effect on the color and geometry of the output image. The image in the middle, generated from a green and red checkerboard, contains artifacts that resemble the checkerboard pattern of the initial image. It also shows a hot air balloon with green patches while the other images do not have any green in them. The image on the right, generated from a black and white checkerboard, is much muted in color compared to the other output images.

These experiments show a way of controlling the final output of an AM GAN by providing it different initial images. By some level of control over the AM GAN, humans can have input regarding the outcome of what the AM GAN generates. This allows for collaboration between the human programmer and the computer on which the AM GAN runs.

In a separate experiment, I wanted to see how objects in an initial image could influence the output. I allowed the GAN to run for 800 iterations and changed the target neuron after 200 iterations. Specifically, the ordering of the target neurons was lemon, balloon, broom, and lemon, again. While I did not start with a specified initial image, the 200th iteration of one target neuron can be interpreted as the initial image of the next target neuron.

Figure 3: The output image at every 200th iteration of the lemon to balloon to broom to lemon sequence.

As shown in Figure 3, the “previous” target neuron does affect the output of the “current” target neuron. The triangular geometric composition of the first lemon image persists in the hot air balloon image. Also, the colors in the hot air balloon image remain present in the broom image. The two different types of lemon images — the lemon cross section in the first image and the lemon fruit on a tree in the last image — show that AM generates a local maximum instead of an absolute maximum. The lemon fruit image is geometrically similar to the broom image while the lemon cross section image is noise-like, not only reinforcing the idea that an initial image can have an impact on the geometric composition of the output image but also showing that an initial image can affect the subtype of canonical image generated.

The sequence of images generated from swapping out the target neuron after a certain number of iterations show yet another way that AM GANs can be used to visualize abstract objects. For example, the broom image shows a broom with qualities of a hot air balloon. This visualization of a combination of objects is different from that of AM using two target neurons simultaneously since previous target neuron's presence is secondary to that of the current target neuron.

Duration of Iterations

While the number of iterations the GAN ran for was never independently explored on its own, its effects can still be observed in the other experiments I conducted. Throughout all experiments, it was evident that more iterations meant more of a convergence towards a canonical image. Usually, after around 250 iterations, it becomes clear in the image generated what the object represented by the target neuron is, and the images generated in subsequent iterations are not that much different from each other.

While images generated in earlier iterations do not necessarily resemble the object represented by the target neuron, they are aesthetically interesting. This can be seen in the experiment where the target neuron was swapped out for another after a certain number of iterations. During the iterations soon after the target neuron was swapped out, the image generated would indicate a sort of breaking away from the previous object represented, its colors and composition start disappearing but are still evident. At the same time, the object specified by the new target neuron begins emerging but is not fully present in the generated image. I call the images generated during the early iterations after the target neuron has been swapped out "transitional images."

Transitional images are able to abstractly evoke imagery of both the "previous" and "current" target neurons. These are different from images generated from activating multiple target neurons simultaneously, which show the objects represented by those target neurons more clearly. Transitional images also differ from images generated in later iterations after a target neuron has been swapped out in which the object represented by the “current” target neuron is distinctly shown and the “previous” target neuron is abstracted if present.

Figure 4: The output image 10 iterations after the target neuron was swapped out from broom to lemon.

Figure 4 shows a transitional image between the broom and lemon neuron. Remnants of broom imagery can be seen in the triangular composition of the image while the emerging lemon imagery can be seen in the yellow, vaguely lemon-shaped object in the center of the image. Both objects are present but are clearly seen.

CONCLUSION

By using an activation maximization generative adversarial network to generate images, I was able to address criticisms of New Aesthetic. Since AM GANs are new and technologically advanced, using an AM GAN addresses Bruce Sterling's criticism that New Aesthetic does not take advantage of recent digital technology. Furthermore, AM GANs can be controlled so that people can have an active role in creating computer-generated images as opposed to finding images. The fact that GANs are a form of artificial intelligence allow for art that fit into Object Oriented Ontology, thus addressing Ian Bogost's criticism of the lack of OOO representation in New Aesthetic.

Another one of my goals was to find ways in which humans can take an active role in the collaboration with the AM GAN to create images. These ways include selecting which and how many neurons are chosen to be target neurons, the initial image used to begin activation maximization, and the number of iterations the GAN goes through. Using these techniques, I was able to generate images that visualized combinations of objects that humans typically have a hard time imagining. These forms of activation maximization allow us to access the "subconscious" of an artificial intelligence since combinations of objects were never directly learned by image recognition neural networks. The resulting images parallel Surrealist paintings, which depict the human subconscious.

Further work related to this thesis include seeing how modifying output images in between iterations would affect later iterations and creating art by juxtaposing AM GAN-generated images with photographs. I hope to continue exploring how humans and computers can collaborate to create art.

WORKS CITED

Anh Nguyen, Alexey Dosovitskiy, Jason Yosinski, Thomas Brox, and Jeff Clune. "Synthesizing the preferred inputs for neurons in neural networks via deep generator networks." May 30, 2016. ArXiv.

Bogost, Ian. "The New Aesthetic Needs to Get Weirder." The Atlantic. April 13, 2012.

Bridle, James. "Sxaesthetic." Booktwo. March 15, 2012. Accessed May 15, 2017.

Cohen, Harold, Becky Cohen, and Penny Nii. *The First Artificial Intelligence Coloring Book*. Los Altos, CA: W. Kaufmann, 1984.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. “Generative Adversarial Networks.” June 10, 2014. ArXiv.

Morton, Timothy. *Hyperobjects*. Minneapolis: University of Minnesota Press, 2014.

Sterling, Bruce. "An Essay on the New Aesthetic." Wired. April 02, 2012.

Warzel, Charlie. "Satan's Credit Card." BuzzFeed. May 21, 2016.

BIOGRAPHY

Angela May Xie was born in Glasgow, Scotland on May 28, 1996. She lived in Witney in Oxfordshire, England and Hefei, China before moving to The Woodlands, Texas in 2001. In Fall of 2014, Angela enrolled at the University of Texas at Austin, majoring in Plan II Honors and Computer Science, and was a student in the Turing Scholars Honors Program. While attending university, Angela became interested in philosophy and software engineering while also maintaining a lifelong interest in visual art; she was awarded second place in the 2016 Signature Course Information Literacy Award for her painting project "The Art of Somatopia," was part of UT's ACM-ICPC competitive programming team, and completed software engineering internships at IBM and Google. Angela graduated in May 2017 and will work for Google in Mountain View, California as a software engineer.